IOP Conf. Series: Materials Science and Engineering **394** (2018) 042058 doi:10.1088/1757-899X/394/4/042058

A Practical Scheme of Smart Grid Privacy Protection

Hanchun Chen[a], Yongjie Yang[*]

School of electronic information, Nantong University, Nantong, China

*Corresponding author e-mail: yang.yj@ntu.edu.com, [a]758191768@qq.com

Abstract. In view of the shortcomings of security loopholes, high overhead and single data types in the current smart grid privacy protection scheme, a practical smart grid privacy protection scheme is proposed in this paper. In terms of security, the combination of an improved Pilliar algorithm and an identity-based pseudonym algorithm realizes a combination of data aggregation technology and identity anonymity technology, making the user's privacy more secure. In the aspect of system overhead, an identity based batch signature verification algorithm is adopted, which makes the number of bilinear pairings calculated only 3 times and has nothing to do with the number of users. Moreover, the improved Paillier algorithm used also greatly reduces the computational overhead. In terms of data types, the combination of super-incremental digital sequences and ladder power consumption fusion algorithms allows the control center to obtain a variety of data. Therefore, this scheme has the characteristics of security, high efficiency and diversification of data types.

1. Introduction

Smart grid is a new generation of next-generation power grid that combines traditional grids with modern information and control technologies. It realizes functions such as balancing loads, dynamically adjusting prices, adjusting power generation plans, and optimizing dispatching of power resources through two-way communications between users and control center. This ensures that the grid operates reliably, efficiently, economically, and sustainably. At the same time, users can not only get a stable power supply, but also save electricity charges by adjusting the power consumption time to the time of low price. However, smart grid also brings users the risk of leakage of privacy. Attackers can spy on users' privacy by eavesdropping data from users' communication with control centers. For example, during the day, the user's power consumption is zero or rarely indicates that the user is not at home, and the attacker can steal the user while he is away. The attacker can also judge the type of the appliance from electricity consumption, and then deduce the user's active state and duration, and monitor the user according to the information, which brings threats to the personal safety of the user.

The data aggregation technology based on Pilliar algorithm [1] is the most commonly used technology of the current smart grid privacy protection scheme. Documents [2-5] adopt a centralized data collection method. The local gateway collects all the electricity consumption data encrypted by the Pilliar algorithm and then aggregates them so that the control center only obtains the total electricity consumption data of the area, thereby protecting users' privacy. In documents [6, 7], each electricity consumption data encrypted by the Pilliar algorithm is aggregated from the leaf node to the root node according to the tree structure path. Finally, the collector sends the cipher text of the total electricity data to the control center, so that the control center also only obtains the total power data of the region.

However, because the private key of the control center can decrypt the encrypted cipher text as well as the single cipher text, this technology has a security hole. Identity anonymity technology is also commonly used in current smart grid privacy protection solutions. Document [8] divides the power consumption data into high frequency data and low frequency data, and uses pseudonyms to send high frequency data, thus protecting the privacy of users. Document [9] uses the group signature technology based on bilinear pairings to achieve the purpose of protecting user privacy. Document [10] uses ring signature technology to construct a privacy protection scheme. Although identity anonymity technology can hide users' identities, it cannot exclude attackers with strong attack capabilities who can crack users' identities.

Because the data aggregation technology based on Pilliar algorithm and the identity anonymity technology (except pseudonym technology) have more exponential operations, their computational complexity is higher, which brings greater computational overhead to the system. In addition, the current signature technology usually does not have the function of batch verification except the BLS short signature [11], which brings a large computational overhead to the system. Even when the BLS short signature is used for batch verification, the number of calculated bilinear pairs still has n+1 times (n is the number of users), and it increases drastically with n. Therefore, the current smart grid privacy protection scheme is inefficient.

In addition, the current smart grid privacy protection scheme only provides the control center with power consumption or power demand, which is too single. It is difficult for the control center to accurately monitor and schedule the power grid based on only one kind of data.

In order to solve the above problems, this paper proposes a practical smart grid privacy protection scheme. The main advantages of this scheme are: Firstly, combining the improved Pilliar algorithm with the identity-based pseudonym algorithm, the combination of data aggregation technology and identity anonymity technology is realized, making users' privacy more secure. Secondly, an identity-based batch signature verification algorithm is adopted, so that the number of calculations of the bilinear pairing when the signature is verified is only 3 times and is independent of the number of users, thereby greatly reducing the computational overhead. Moreover, the improved Paillier algorithm used also greatly reduces the computational overhead. Thirdly, in terms of data types, the combination of the super-incremental digital sequence and the ladder power consumption fusion algorithm enables the control center not only to obtain the total power consumption of the users in a certain area, but also to obtain the total power consumption demand and the ladder power consumption and the corresponding number of users.

2. System model, security requirement and design goal

2.1. System model

The model of smart grid communication system presented in this paper is shown in Figure 1. The model consists of four entities: smart meter (SM), gateway (GW), control center (CC), and trusted third party (TTP). As can be seen from the figure, the CC manages k areas, each area has one GW and n users U= {U_1, U_2, ⋯, U_n}, and the SM is installed in the user room. Because each SM is very close to the local GW, they are connected by cheap wireless communication such as WiFi. The distance between the GW and the CC is usually relatively long, so they are connected by high-bandwidth, low-latency wired communication such as optical fiber.

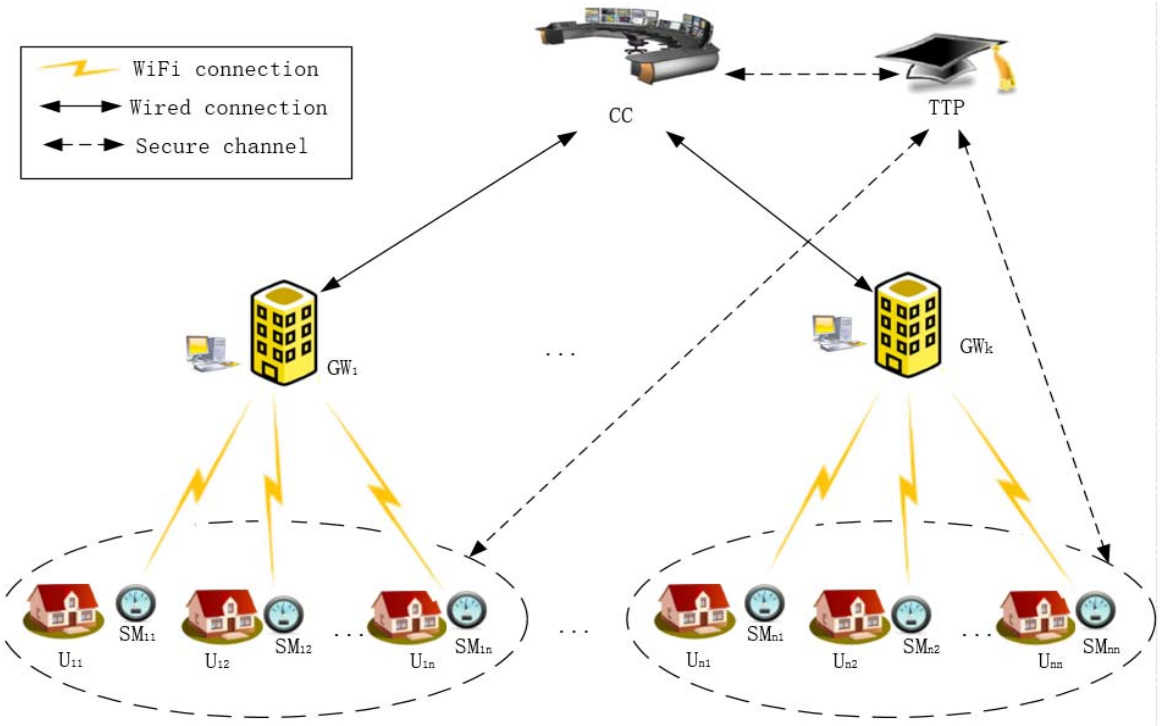

Figure 1. System model.

• SM: Send encrypted power data to the GW on a regular period (usually 15 minutes). Each SM_i has a built-in anti-tamper device. The device is composed of an authentication module, a pseudonym generation module, and a private key generation module for generating parameters such as pseudonym and key. The private key generation module has a built-in key K_i of the AES symmetric encryption algorithm, and SM_i cannot be known, only the GW knows the key.

• GW: Responsible for authenticating and aggregating data from users and forwarding the aggregated data to the CC. In addition, it also broadcasts the data returned by the CC to the SM.

• CC: CC is the core entity of the system and has powerful computing capabilities. Its main functions include initializing the system, collecting, processing, and analysing data forwarded by the GW, and monitoring and scheduling the power grid based on the results of the analysis to achieve optimal conditions.

• TTP: TTP is a fully trusted entity (usually served by a power company) in the smart grid. In this scheme, TTP only participates in the allocation of secret parameters and will be offline thereafter.

2.2. Security requirement

In the security model of this paper, we assume that SM is honest and trustworthy; TTP is completely trustworthy; GW and CC are honest and curious, that is, they can comply with the protocol to process the received data and do not leak the data, but to spy on the user's power consumption information as much as possible in the execution of the protocol. In addition to passive attacks such as eavesdropping on user data, external attackers may also perform active attacks such as tampering or replacement. Based on the above analysis, the security requirements for this scheme are as follows:

(1) Confidentiality. Attackers can't obtain any user's power consumption information by eavesdropping on wireless or wired channels, even by invading the gateway or the control center's database.

(2) Authentication and integrity. In order to prevent malicious attackers from posing as legitimate users to send data, GW and CC must have the function of identity authentication. At the same time, in order to prevent the attacker from forging or altering the user's power consumption data, the GW and

the CC must also have the function of verifying the authenticity and integrity of the data to ensure that the data is real data from legitimate users.

(3) Anonymity. In order to ensure that the user's privacy is not leaked from the source, the internal entities GW and CC of the grid cannot obtain the power consumption information of a single user. Even if they know the power consumption information of a certain user, they cannot know their true identities.

2.3. Design goal

The overall design goal of this project is to propose a practical scheme of smart grid privacy protection. The objectives are as follows:

(1) Safety. If the user's privacy is leaked during the operation of the smart grid, it may bring potential threats to user's property and life, thereby hindering the healthy development and rapid promotion of the smart grid. Therefore, ensuring the security of user's privacy is the basic goal of the scheme.

(2) High efficiency. If the scheme does not pay attention to the high efficiency, it will bring unacceptable computational overhead to various entities of the smart grid, and it will also bring serious congestion to the communications network, and even lead to grid collapse. Therefore, the scheme should reduce the cost and improve the efficiency as much as possible so as to reflect the characteristic of the real-time performance of the smart grid.

(3) Variety of data types. To achieve CC's accurate and effective monitoring and scheduling of smart grids, in addition to providing CC with the total power consumption of users in each region, they should also provide various data such as total electricity demand, ladder power consumption, and the corresponding number of users. Therefore, the scheme should meet the CC's need for diversification of data types.

3. Preliminary knowledge

3.1. Bilinear Pairing

Let $\mathbb{G}$ be an additive cyclic group generated by generator P, $\mathbb{G}_T$ is a multiplicative cyclic group, and the order of $\mathbb{G}$ and $\mathbb{G}_T$ is q, where q is a large prime number. When the mapping e: $\mathbb{G} \times \mathbb{G}=\mathbb{G}_T$ satisfies the following three properties, it is called bilinear mapping:

• Bilinearity: $e(\alpha P_1, \beta P_2) = e(P_1, P_2)^{\alpha\beta}$ for $\forall P_1, P_2 \in \mathbb{G}$ and $\forall \alpha, \beta \in \mathbb{Z}_q^*$.

• Non-degeneracy: $e(P, P) \neq 1_{\mathbb{G}_T}$ for $\exists P \in \mathbb{G}$.

• Computability: There exists an effective polynomial time algorithm to calculate the value of bilinear pairing.

Definition 1 (Bilinear Pairing Generation Algorithm): $\mathcal{G}en$ is a bilinear parameter generation function that takes a safety parameter κ as input and outputs a 5-tuple (q, P, $\mathbb{G}$, $\mathbb{G}_T$, e).

3.2. Improved Paillier algorithm

The algorithm consists of three parts: key generation, encryption and decryption [12-14].

(1) Key generation: Select a security parameter τ and two large prime numbers $p'$and $q'$of lengthτ, calculate $N=p'q'$ and λ=lcm (p'-1, q'-1) and $b=\lambda^{-1}$ mod N. Define the function $L(u) = (u-1)/N$, then the public key is $pk=N$ and the private key is $sk=(\lambda, b)$.

(2) Encryption: Randomly select multiple $r \in \mathbb{Z}_n^*$ before encryption, and pre-calculate multiple $s=r^N \bmod N^2$. Given a plaintext message $m \in \mathbb{Z}_N$, randomly select one s and calculate the cipher text $c=E(m) = (1+mN) \cdot s$.

(3) Decryption: Given the cipher text $c \in \mathbb{Z}_{N^2}^*$, restore the corresponding plaintext message as $m=D(c) = L(c^{\lambda} \bmod N^2) \cdot b \bmod N$.

According to [12, 13], the improved Paillier algorithm has the same semantic security features and additive homomorphism without reducing the security of the traditional Paillier algorithm. Compared with the traditional Paillier algorithm, the improved Paillier algorithm has one less key exponent for private key generation and encryption.

3.3. Scheme of identity-based pseudonym generation and batch signature verification

The scheme includes three parts: system initialization, pseudonym generation algorithm and batch verification algorithm [15]. The following are the details of it.

(1) System initialization

Each user (U) in the program is embedded with a tamper-resistant device. The device is composed of an authentication module, a pseudonym generation module, and a private key generation module. In the private key generation module, there is a key *K* of the built-in AES symmetric encryption algorithm, and U cannot be known. Only the trusted authority (TA) knows the key.

Let $\mathbb{G}$ be an additive cyclic group generated by generator *P*, $\mathbb{G}_T$ is a multiplicative cyclic group, and the order of $\mathbb{G}$ and $\mathbb{G}_T$ is *q*, where q is a large prime number. Let *me*: $\mathbb{G} \times \mathbb{G}=\mathbb{G}_T$ be a bilinear map. Select the AES symmetric encryption algorithm *E* () and two hash functions:

$H(\cdot):\{0,1\}^* \to \mathbb{G}$, $h(\cdot):\{0,1\}^* \to \mathbb{Z}_q^*$.

TA randomly selects$s_1, s_2 \in \mathbb{Z}_q^*$ as the master key $s=(s_1,s_2)$, and calculates $P_{pub1}=s_1P$, $P_{pub2}=s_2P$ as the public key $P_{pub}=(P_{pub1}, P_{pub2})$.

(2) Pseudonym generation algorithm

1) U filed a registration application with TA in its true identity *ID*.

2) After TA verifying the *ID* of U, it first encrypts its private key *s* with the symmetric key *K* of the U's tamper proof device and encrypts the $E_K(s)$, and then sends the network admission certificate Permit containing $E_K(s)$ to the U through the secure channel.

3) U enter the *Permit* and its own *ID* into the tamper-resistant device to start the device. After the authentication module successfully verifies the *Permit* and *ID*, the $E_K(s)$ and *I*D are transmitted to the pseudonym generating module.

4) The pseudonym generation module selects a random number $r\in \mathbb{Z}_q^*$, calculates $PID_1=rP$ and $PID_2=ID\oplus H(rP_{pub1})$, and obtains pseudonym $PID=(PID_1, PID_2)$, and transfer the *PID* and $E_K(s)$ to the private key generation module.

(3) Batch signature verification algorithm

1) The private key generation module decrypts $E_K(s)$ to obtain $s=(s_1,s_2)$, and calculates $SK_1=s_1PID_1$ and $SK_2=s_2H\ (PID_1 \parallel PID_2)$ based on pseudonym $PID=(PID_1, PID_2)$, gets the private key $SK=(SK_1, SK_2)$.

2) If there are *n* users, denoted by U_i (i=1, 2,⋯ n) respectively, U_i uses the private key $SK_i=(SK_{i1}, SK_{i2})$ calculate the signature of the message M_i $\sigma_i=SK_{i1}+h\ (M_i)\ SK_{i2}$ and send the final data packet $<PID_i, M_i, \sigma_i>$ to the TA.

3) After receiving the data packet sent by U_i(i=1,2,⋯,n) $<PID_1, M_1, \sigma_1>$, $<PID_2, M_2, \sigma_2>$, ⋯, $< PID_n , M_n , \sigma_n >$, TA can perform batch verification, that is, verify the equation $e(\sum_{i=1}^{n}\sigma_i,P)=e(\sum_{i=1}^{n}(PID_1 + h(M_i)\ H(PID_{i1} \parallel PID_{i2})), P_{pub})$ is established. If it is established, the signature is valid, otherwise the signature is invalid.

Obviously, the number of calculations of the bilinear map e (·, ·) is reduced from 3*n* times for single signature verification to 3 times for batch verification, and is independent of the number of users.

3.4. BLS short signature

BLS short signature consists of three parts: key generation, signature and verification [16].

Let $\mathbb{G}$ be an additive cyclic group of order prime *q*, $\mathbb{G}_T$ be a cyclic multiplicative group of the same order, *P* be a generator of$\mathbb{G}$, and mapping *e*: $\mathbb{G} \times \mathbb{G} \to \mathbb{G}_T$ be a bilinear map, *H*: $\{0,1\}^* \to \mathbb{G}$ is a hash function.

(1) Key generation: Select $x\in \mathbb{Z}_q$ as the private key and calculate *Y*=*xP* as the public key.

(2) Signature: Given the message $m\in \{0,1\}^*$, calculate the signature σ=*xH* (*m*).

(3) Verification: Verify that the equation *e* (σ, *P*) =*e* (*H* (*m*), *Y*) holds. If it is established, accept m, otherwise reject m.

4. Scheme realization

This paper is composed of system initialization, user data collection, data aggregation and data processing. In order to facilitate analysis, we take any region as an example. Therefore, neither GW nor SM in the solution is marked with a subscript indicating a specific region.

4.1. System initialization

(1) System parameters and master key generation

1) CC selects a security parameter and generates parameters (q,P,$\mathbb{G}$, $\mathbb{G}_T$,e) by calling function $\mathcal{G}en(\mathcal{K})$.

2) CC selects a security parameterτ, calculates the public key pk=N of the improved Paillier algorithm and the private key sk= (λ, b).

3) CC selects the random number $x\in \mathbb{Z}_q^*$ as the master key, and calculates Y=xP as the public key of the signature.

4) CC selects two hash functions: $H(\cdot):\{0,1\}^* \rightarrow \mathbb{G}$, $h(\cdot):\{0,1\}^* \rightarrow \mathbb{Z}_q^*$.

5) CC publishes system public parameters $param$={ q, P, $\mathbb{G}$, $\mathbb{G}_T$,e,Y,N,H,h,} and secretly stores the master key (x, λ,b).

(2) System entity registration and pseudonym generation

1) GW selects the random numbers x_{GW1}、$x_{GW2} \in \mathbb{Z}_q^*$ as the private key of the signature x_{GW}=(x_{GW1},x_{GW2}), and calculates Y_{GW1}=x_{GW1}P and Y_{GW2}=x_{GW2}P as the public key Y_{GW}=(Y_{GW1},Y_{GW2}) of the signature.

2) GW sends a registration request to CC with its own real identityID_{GW}. After verifying GW'sID_{GW}, CC sends the network access certificate *Cert* to GW through the secure channel.

3) The user U_i(i=1, 2,⋯ n) sends a registration request to the GW with its own real identity ID_i of theSM_i. After verifying ID_i ofSM_i, GW first encrypts its own private key x_{GW} with symmetric key K_i of the built-in tamper-resistant device of SM_i to obtain $E_{K_i}(x_{GW})$, and then sends the network certificate $Permit_i$ with $E_{K_i}(x_{GW})$ to SM_i through a secure channel.

4) SM_i Inputs $Permit_i$ and it's ID_i into the built-in tamper-resistant device to start the operation of each module in the device.

① After verifying $Permit_i$ and ID_i, the authentication module transfers ID_i and $E_{K_i}(x_{GW})$ to the pseudonym generation module.

② The pseudonym generation module selects a random number $r_i' \in \mathbb{Z}_q^*$, calculates PID_{i1}=r_i'P and PID_{i2}=ID_i⊕ H($r_i' Y_{GW1}$) obtains pseudonym PID_i=(PID_{i1}, PID_{i2}), and then transfers PID_i and $E_{K_i}(x_{GW})$ to the private key generation module.

③The private key generation module decrypts $E_{K_i}(x_{GW})$ to obtain x_{GW}= (x_{GW1},x_{GW2}), and then calculates X_{i1} =$x_{GW1} PID_{i1}$ and X_{i2}=$x_{GW2} H$ ($PID_{i1} \parallel PID_{i2}$) according to the pseudonym PID_i= (PID_{i1}, PID_{i2}),and the private key X_i= (X_{i1}, X_{i2}) is obtained.

(3) Secret parameter assignment

1) Assuming that the maximum number of users in each area is not greater than n, a total of l+1 types of electricity data (T_1,T_2,⋯,T_l,T_{l+1}) are reported in the smart grid communication, and each T_i The value is less than the constant d. TTP selects a super-incrementing digital sequence $\vec{a}$=(a_1=1,a_2,⋯,a_l, a_{l+1}), where a_2,⋯,a_l, a_{l+1} are large prime numbers. And the length of a_i $|a_i| \geq \mathcal{K}$,$\sum_{j=1}^{i} a_j \cdot n \cdot d < a_i$ (where i=2, ⋯, l), $\sum_{i=1}^{l+1} a_i \cdot n \cdot d <$$N$.

2) In order to facilitate the calculation of the step power, TTP selects a large prime number $\mathcal{B}$ which is much larger than the power consumption of any user in the area within 15 minutes.

3) The TTP sends the super-incrementing digital sequence $\vec{a}$ and the large prime $\mathcal{B}$ through the secure channel to the CC and all the SM_i in the area.

4.2. User data collection

(1) Composition of user data

The data sent by each user U_i(i=1, 2,⋯, n) through SM_i in this scheme includes not only the power consumption m_i, but also the power demand m'_i and the ladder power consumption. The following is an expression of the ladder power consumption.

Assume that a total of l step powers (Θ_1,Θ_2,⋯,Θ_l) are used to measure the power consumption of each user, and the userU_i power consumption $m_i \in[\Theta_k,\Theta_{k+1}]$. If Δm_{ij}(j=1, 2,⋯, l) are used to represent the power consumption of m_i on each ladder, then there are:

When 1<k≤l,

$$\Delta m_{ij}=\begin{cases}\Theta_1, & j=1;\\ \Theta_j-\Theta_{j-1}, & 1<j\le k;\\ m_i-\Theta_k, & k<j\le k+1;\\ 0, & k+1<j\le l.\end{cases} \tag{4-1}$$

When k=1,

$$\Delta m_{ij}=\begin{cases}m_i, & j=1;\\ 0, & 1<j\le l.\end{cases} \tag{4-2}$$

In order to facilitate the recovery of the received data by the control center, this paper transforms Δm_{ij} into another form:

$$\Delta' m_{ij}=\begin{cases}\Delta m_{ij}, & 1\le j\le k, k+1<j\le l;\\ \Delta m_{ij}+\mathcal{B}, & j=k+1.\end{cases} \tag{4-3}$$

(2) User's report generation

1) Encryption:SM_i randomly selects a pre-calculated s_i=r_i^N modN^2, then combines the super-incrementing digital sequence$\vec{a}$, the ladder power consumption $\Delta' m_{ij}$ and the power demand $\Delta' m_{ij}$ for improved Paillier algorithm encryption:

$$c_i=[1+(a_1\Delta' m_{i1}+a_2\Delta' m_{i2}+\cdots+a_l\Delta' m_{il}+a_{l+1}m'_i)\ \mathrm{N}]\cdot s_i \tag{4-4}$$

2) Signature: SM_i uses the private key $X_i=(X_{i,1},X_{i,2})$ to use identity-based batch signature verification algorithm for ciphertext c_i to calculate the signature:

$$\sigma_i=X_{i1}+\mathrm{h}\,(c_i)\,X_{i2} \tag{4-5}$$

3) Report: SM_i send the data packet $C \parallel \sigma_{GW} \parallel ID_{GW} \parallel ID_{CC} \parallel TS$ to the local GW, where TS is the current timestamp.

4.3. Data Aggregation

After the local GW receives the data packet $c_i \parallel \sigma_i \parallel PID_i \parallel$TS (i=1, 2,⋯, n) sent by the n SMs, it performs the following operations:

(1) Batch signature verification: GW first calculates h (c_i) and H ($PID_{i1} \parallel PID_{i2}$) ($i$=1, 2,⋯,$n$), and then verifies the equation:

$$\mathrm{e}(\textstyle\sum_{i=1}^{n}\sigma_i,\mathrm{P})=\mathrm{e}(\sum_{i=1}^{n}PID_{i1},Y_{GW1})\mathrm{e}(\sum_{i=1}^{n}h(c_i)H(PID_{i1}\parallel PID_{i2}),Y_{GW2}) \tag{4-6}$$

Whether or not it is established. If it is established, GW accepts these packets, otherwise GW will discard these packets.

(2) Data aggregation:

$$\mathrm{C}=\prod_{i=1}^{n} c_i \bmod N^2 \tag{4-7}$$

(3) Signature: GW uses private key x_{GW1} to compute the signature of aggregated ciphertext *C*:

$$\sigma_{GW}=x_{GW1}\mathrm{H}(\mathrm{C}\| ID_{GW} \| ID_{CC} \|\mathrm{TS}) \tag{4-8}$$

Where *TS* is the current timestamp.

(4) Report: GW sends the data packet $C \| \sigma_{GW} \| ID_{GW} \| ID_{CC} \| TS$ to CC, where *TS* is the current timestamp.

4.4. Data Processing

After receiving the data packet from GW, CC performs the following operations:

(1) Signature verification: CC verifies the equation:

$$\mathrm{e}(\sigma_{GW},\mathrm{P})=\mathrm{e}(\mathrm{H}(\mathrm{C}\| ID_{GW} \| ID_{CC} \|\mathrm{TS}),Y_{GW1}) \tag{4-9}$$

Whether or not it is established. If it is established, CC accepts the packet, otherwise CC will discard the packet.

(2) Decryption of aggregated cipher text: because

$$\begin{aligned}\mathrm{C}&=\prod_{i=1}^{n} c_i \bmod N^2\\ &=\prod_{i=1}^{n}[1+(a_1\Delta' m_{i1}+a_2\Delta' m_{i2}+\cdots+a_l\Delta' m_{il}+a_{l+1}m'_i)N]\cdot s_i \bmod N^2\\ &=[1+\mathrm{N}\textstyle\sum_{i=1}^{n}(a_1\Delta' m_{i1}+a_2\Delta' m_{i2}+\cdots+a_l\Delta' m_{il}+a_{l+1}m'_i)]\cdot\prod_{i=1}^{n} s_i \bmod N^2\\ &=[1+\mathrm{N}\,(a_1\textstyle\sum_{i=1}^{n}\Delta' m_{i1}+a_2\sum_{i=1}^{n}\Delta' m_{i2}+\cdots+a_l\sum_{i=1}^{n}\Delta' m_{il}+a_{l+1}\sum_{i=1}^{n}m'_i)]\cdot\prod_{i=1}^{n} s_i \bmod N^2\end{aligned} \tag{4-10}$$

Let $\bar{M}=a_1\sum_{i=1}^{n}\Delta' m_{i1}+a_2\sum_{i=1}^{n}\Delta' m_{i2}+\cdots+a_l\sum_{i=1}^{n}\Delta' m_{il}+a_{l+1}\sum_{i=1}^{n}m'_i$, $S=\prod_{i=1}^{n} s_i$, Then $C=(1+\bar{M}N)\cdot S$ is still the ciphertext form of the improved Paillier algorithm. Therefore, CC uses the private key (λ, b) to decrypt *C*:

$$\bar{M}=(a_1\textstyle\sum_{i=1}^{n}\Delta' m_{i1}+a_2\sum_{i=1}^{n}\Delta' m_{i2}+\cdots+a_l\sum_{i=1}^{n}\Delta' m_{il}+a_{l+1}\sum_{i=1}^{n}m'_i) \bmod \mathrm{N} \tag{4-11}$$

(4) Recovery of data

1) Recovery of aggregated data

In formula (4-11), let $U_j=\Delta' M_j=\sum_{i=1}^{n}\Delta' m_{ij}$(j=1,2,⋯,l),$U_{l+1}=M'=\sum_{i=1}^{n}m'_i$, then they are aggregated data to be restored. The CC can recover these data through the Algorithm 1.

Algorithm 1:

Iutput:$\bar{M}$ and $\vec{a}=(a_1=1, a_2, \cdots, a_l, a_{l+1})$
Output:$(U_1,\ U_2,\ \cdots, U_l,\ U_{l+1})$
  Set $V_{l+1}=U_{l+1}$,
  For *j*=*l*+1to 2 do
    $V_{j-1}=V_j \bmod a_j$
    $U_j=(V_j-V_{j-1})/a_j$
  End for
  $U_1=V_1$
Return $(U_1,\ U_2,\ \cdots, U_l,\ U_{l+1})$
End procedure

$(U_1, U_2, \cdots, U_l, U_{l+1})$ are $(\Delta' M_1, \Delta' M_2, \cdots, \Delta' M_l, M')$.

2) Recovery of power consumption data

The total power consumption of each ladder is ΔM_j=$\sum_{i=1}^{n} \Delta m_{ij}$.CC can continue recovering the power consumption of each ladder ΔM_j and the corresponding number of users n_j by Algorithm 2.

Algorithm 2:

Iutput:$(\Delta' M_1, \Delta' M_2, \cdots, \Delta' M_l)$
Output :$(\Delta M_1, \Delta M_2, \cdots, \Delta M_l)$ and $(n_1, n_2, \cdots, n_l)$
For j =1to l do
ΔM_j=$\Delta' M_j$ mod $\mathcal{B}$
n_j=$(\Delta' M_j - \Delta M_j)/\mathcal{B}$
End for
Return $(\Delta M_1, \Delta M_2, \cdots, \Delta M_l)$ and $(n_1, n_2, \cdots, n_l)$
End procedure

In Algorithm 2, because $\Delta' M_j$=ΔM_j+ n_j $\mathcal{B}$(where $\Delta' M_j$=$\sum_{i=1}^{n} \Delta' m_{ij}$, ΔM_j=$\sum_{i=1}^{n} \Delta m_{ij}$), ΔM_j=$\Delta' M_j$ mod$\mathcal{B}$, n_j=$(\Delta' M_j - \Delta M_j)/\mathcal{B}$.CC calculates the total power consumption M=$\sum_{j=1}^{l} \Delta M_j$. Finally, CC gets the total power consumption M in the area, and the total power demandM', each step power consumption ΔM_j(j=1, 2, ⋯, l) and the corresponding number of users n_j.

5. Security analysis

This scheme satisfies all the security requirements presented in 2.2. The specific analysis is as follows:

(1) Confidentiality

Each user's power consumption data is first encrypted by the SM using the improved Paillier algorithm to generate cipher text, and then sent to the local GW through the wireless channel; GW aggregates all the collected cipher texts and forwards them to the CC through the cable channel. During the communication process, electricity data is transmitted in ciphertext. Because the improved Paillier algorithm is semantically secure under selective plaintext attacks, the cipher text is semantically secure. Therefore, although an external attacker can obtain cipher text by eavesdropping on the wireless channel, or the internal entity GW can directly receive the cipher text, none of them can obtain any user's power consumption information without the private key. So this scheme can ensure the confidentiality of all users' power consumption data.

(2) Anonymity

1) Anonymity of electricity data: Although the internal entity CC has a private key, the cipher text it obtains is a cipher text after GW's aggregating. So CC obtains the power consumption data after it decryptes is the total power consumption data of the area. CC cannot get any user's power consumption data by analyzing the total power consumption data, thus realizing the anonymity of all users' power data.

2) Anonymity of user identity: We also consider the following cases: Assuming that all or part of the cipher text sent by SMs is sent directly to CC for various reasons. Since the private key of the CC can decrypt both the cipher text of the aggregated data and the cipher text of the single data, once the CC obtains the cipher text of these individual users, the private key can be used for decryption by the CC, thereby obtaining the detailed power consumption data of these users. In addition, external attackers with strong attack capabilities can obtain this data by invading CC's database. In order to solve this problem, the identity-based pseudonym algorithm adopted in this scheme can achieve the anonymity of the users' identities, which is equivalent to adding an “insurance” to the users' privacy. A series of pseudonyms generated by each SM through the algorithm makes CC and even the GW and the external attackers unable to associate with the corresponding user identity even if they obtain detailed power consumption data, ensuring the privacy of all users is more secure.

(3) Authenticity and integrity

In this scheme, each user's SM must use his private key to sign the cipher text of the power data

according to the identity-based batch signature verification algorithm before sending the report, so the local GW should use its own public key to verify the signature after receiving the users' reports. Similarly, before the GW sends a report, it uses its own private key to make BLS short signature on the aggregated cipher text, so CC uses GW's public key to verify the signature after receiving the GW's report. Based on the difficulty of CDH problem, the identity based batch verification signature and BLS short signature can resist the selective message forgery attack under the random oracle model (ROM). Therefore, the scheme is authentic for messages and ensures the reliability of the source. On the other hand, if an external attacker falsifies or falsifies power usage data, GW and CC will fail to perform signature verification, thereby rejecting the forged power consumption data, thereby ensuring the integrity of the power consumption data.

6. Performance analysis

We evaluate the performance of this solution from three aspects: computational overhead, communication overhead and function.

6.1. Computational overhead

In order to facilitate analysis, we respectively use C_e, C_m, C_{et} and C_p to represent the calculation overhead of an exponent operation on $\mathbb{Z}_{N^2}^*$, the calculation overhead of a multiplication operation on $\mathbb{G}$, the calculation overhead of an exponent operation and a single bilinear operation on $\mathbb{G}_T$. Compared with the above operations, the computational overhead of other operations is negligible. The literature [16] on the computer of Intel Core i5-2430, 2.4GHz CPU, 2GB RAM, uses the MIRACL library [17] to carry on the operation time overhead experiment, obtainsC_e, C_m and C_p respectively takes 9.78ms, 1.18ms and 22.84ms.

In our scheme, each SM_i needs to perform 1 exponential operation on $\mathbb{Z}_{N^2}^*$ to generate the cipher text C_i and 1 multiplicative operation on $\mathbb{G}$ to generate the signature σ_i, so the total computational overhead of the SM_i is C_e+C_m. The local GW first performs batch signature verification after collecting n users' data packets, this operation requires 3 bilinear operations. Then data aggregation is performed, which requires n-1 multiplication operation on $\mathbb{Z}_{N^2}^*$,but compared with the cost of the exponential operation on $\mathbb{Z}_{N^2}^*$,the overhead of multiplication on $\mathbb{Z}_{N^2}^*$ can be ignored, so the computational overhead of data aggregation can be ignored. Finally, 1 multiplication operation on $\mathbb{G}$ is used to generate the signatureσ_{GW}, so the total computation overhead of GW is 3 C_p+C_m.After receiving a packet sent by GW, CC first needs to do signature verification, which requires 2 bilinear operations.Then the improved Paillier decryption algorithm is performed, which requires 1 exponent operation on$\mathbb{Z}_{N^2}^*$, so the total computational overhead of CC is $2C_p$+C_e.

Table 1 shows the computational overhead of our scheme and other three schemes, including the main computational overhead of SM, local GW and CC, and the total calculation time of the whole system. The TRAD scheme refers to the traditional scheme. That is, each SM uses the traditional Paillier algorithm to encrypt one-dimensional data and uses the BLS short signature, and GW verifies all SMs' signatures one by one and does not aggregate the cipher text.

In the table, l refers to the dimension of the data (l≥2), and n refers to the number of users in the area (n≫10).

Table 1. Computational overhead comparison.

Scheme	Computational overhead			Total calculation time (ms)
	SM	GW	CC	
TRAD	$2l$ C_e+C_m	$2nC_p$+C_m	$2C_p$+lC_e	48.84+29.34l+45.68n
EPPA[2]	(l+1) C_e+C_m	(n+1) C_p+C_m	$2C_p$+C_e	90.44+9.78l+22.84n
EPPDR[3]	$2C_e$+$2C_m$	3n C_p+$2C_m$	$3C_p$+C_e	102.58+68.52n
Our scheme	C_e+C_m	3 C_p+C_m	$2C_p$+C_e	136.12

As can be seen from table 1, although our scheme contains *l*+1-dimensional power consumption data, the computational overhead of SM, GW and CC are all less than or equal to the other three schemes, making the total calculation time much smaller than the other three schemes, and they do not vary with the changes of *l* and *n*.

6.2. Communication overhead

According to the rules of cryptographic security, the length of RSA parameter N is 1024bit, and the length of $\mathbb{G}$ in the elliptic curve is 160bit, then the length of Paillier and improved Paillier cipher text on $\mathbb{Z}_{N^2}^*$ is 2048bit, the length of the signature is 160bit.We set the identity (address) of all entities length | ID | and timestamp length | TS | are all 32bit. Usually each SM_i sends packets to the local GW as $c_i \parallel \sigma_i \parallel ID_i \parallel ID_{GW} \parallel$TS, the data packet sent by the GW to the CC is $C \parallel \sigma_{GW} \parallel ID_{GW} \parallel ID_{CC} \parallel$TS.

In our scheme, the communication overhead from each SM_i to the local GW is 2048+160+32+32+32=2304 bit, and the communication overhead from GW to CC is 2048+160+32+32+32=2304 bit. It should be noted that the cipher text c_i of each SM_i in the scenario of this paper contains *l*+1 dimensional data, and the corresponding local GW's aggregated cipher text *C* also contains *l*+1 dimensional data.

Table 2 shows the communication overhead of our scheme and the other three schemes, including the communication overhead from SM to GW and GW to CC respectively. The calculation standard for the communication overhead of other schemes is the communication overhead of the entire data packet in the case of sending *l*+1-dimensional data.

Table 2. Comparison of communication overhead.

Scheme	Communication overhead (bit)	
	SM to GW	GW to CC
TRAD	2048(*l*+1)+256=2304+2048*l*	2048(*l*+1)+256=2304+2048*l*
EPPA[2]	2048*2+256=4352	2048*2+256=4352
EPPDR[3]	2048(*l*+1)+256=2304+2048*l*	2048(*l*+1)+256=2304+2048*l*
Our scheme	2048+256=2304	2048+256=2304

As can be seen from table 2, the communication overhead of our scheme from SM to GW and GW to CC are far less than the other three schemes.

6.3. Function

Table 3 gives a comparison of the functions of our scheme and the other three schemes.

Table 3. Comparison of functions

Scheme	Data types of collection	Signature verification		Types of power consumption data obtained by CC			
		Batch verification	Number of calculation of bilinear pairings	Power comsunption related information	Total power consumption	Total power demand	Step power consumption and number of users
TRAD	One kind		2*n*		√		
EPPA[2]	Variation	√	*n*+1	√	√		
EPPDR[3]	One kind		3*n*			√	
Our scheme	Variation	√	3		√	√	√

As can be seen from Table 3, except that it does not have power comsunption related information function of the EPPA scheme, our scheme has all the listed functions, and the number of bilinear pairs

is only 3.Moreover, power comsunption related information of the EPPA scheme is about the time and purpose of power comsunption, and has no substantive effect on the monitoring and scheduling of the CC.

Based on the above analysis, we can see that compared with the other three solutions, our scheme has a great advantage in terms of computational overhead, communication overhead, and function, which shows that our scheme has good performance.

7. Conclusion

This paper proposes a practical smart grid privacy protection scheme aiming at the defects of current smart grid privacy protection schemes, combined with improved Pilliar algorithm, identity-based pseudonym algorithm and batch signature verification algorithm. Through security analysis and performance analysis, it is proved that our scheme achieves the design goals of security, high efficiency, and diversified data types. Therefore, the conclusion that our scheme has strong practicability is obtained.

Acknowledgments

This work was financially supported by 1st Phase Project of Colleges' Brand Specialty Construction of Jiangsu Province, China (No.PPZY2015B135).

References

[1] P. Paillier. Public-key cryptosystems based on composite degree residuosity classes[C]. Stem J ed. Advances in cryptology - EUROCRYPT '99.Springer Berlin Heidelberg, 1999: 223 - 238.
[2] R.X. Lu, X.H. Ling, X. Li, et al. EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications [J]. IEEE Transactions on Parallel and Distributed System, 2012, 33 (9): 1621 - 1631.
[3] H.W. Li, X.D. Lin, H.M. Yang, et al. EPPDR: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid [J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25 (8): 2053 - 2064.
[4] H. Shen, M.W. Zhang. A privacy protection smart grid multi-level user power aggregation control scheme [J]. Cryptography Journal, 2016, 3 (2):171 - 191.
[5] L. Chen, Y.F. Lin. Smart grid security data fusion technology based on homomorphic encryption [J]. Modern Electronic Technology, 2016, 9 (39): 82 - 86.
[6] F.J. Li, B. Luo. Preserving Data Integrity for Smart Grid Data Aggregation [C]. 2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm).IEEE, 2012: 366 - 371.
[7] P. Deng, L. Yang. A secure and privacy-preserving communication scheme for advanced metering infrastructure [C]. Innovative Smart Grid Technologies, 2012 IEEE PES. IEEE, 2012:1 - 5.
[8] C. Efthymiou, G. Kalogridis. Smart grid privacy via anonymization of smart metering data[C]. 2010 1st IEEE International Conference on Smart Grid Communications. IEEE, 2010: 238 - 243.
[9] S.H.M. Zargar, M.H. Yaghmaee. Privacy preserving via group signature in smart grid [C]. 2013 3st International Conference on Computer and Knowledge Engineering. IEEE, 2013: 368 - 373.
[10] J. Zhao, J.Q. Liu, Z. Qin, et al. Privacy Protection Scheme Based on Remote Anonymous Attestation for Trusted Smart Meters [J]. IEEE Transactions on Smart Grid, 2017, 99: 1 - 8.
[11] D. Boneh, B. Lynn, H.J. Shacham. Short Signatures from the Weil Pairing [J]. Journal of Cryptology. Springer Berlin Heidelberg, 2004: 297 - 319.
[12] F. Yan, W. Xu, Y. Feng, et al. Improved Paillier algorithm design for smart meter privacy protection scheme [J]. Power Information and Communication Technology, 2016, 12 (14): 52 - 57.
[13] W. Xu. Design of scheme for privacy protection of smart meter [D]. Beijing: North China Electric Power University, 2016.

[14] B. Yang, G.Z Xiao. Modern Cryptography (3rd Edition) [M]. Beijing: Tsinghua University Press, 2015.
[15] C.X. Zhang, P.H. Ho, J. Tapolcai. On batch verification with group testing for vehicular communications [C]. International Conference on Wireless Networks. Springer Berlin Heidelberg, 2011: 1851 - 1865.
[16] O.R.M. Boudia, S.M. Senouci, Feham M. Elliptic Curve Based Secure Multidimensional Aggregation for Smart Grid Communications [J]. IEEE Sensors, 2017, 23 (17): 7550 - 7757.
[17] Certivox. Multiprecision integer and rational arithmetic c/c++ library (MIRACL), 2014, https://github.com/miracl/MIRACL.